The solution structure of the Mu Ner protein reveals a helix-turn-helix DNA recognition motif

Teresa E Strzelecka, G Marius Clore* and Angela M Gronenborn*

Laboratory of Chemical Physics, Building 5, National Institutes of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0520, USA

Background: The Mu Ner protein is a small (74 amino acids), basic, DNA-binding protein found in phage Mu. It belongs to a class of proteins, the cro and repressor proteins, that regulate the switch from the lysogenic to the lytic state of the phage life cycle. There is no significant sequence identity between Mu Ner and the cro proteins of other phages, despite their functional similarity. In addition, there is no significant sequence identity with any other DNA-binding proteins, with the exception of Ner from the related phage D108 and the Nlp protein of *Escherichia coli*. As the tertiary structures of Mu Ner and these two related proteins are unknown, it is clear that a three-dimensional (3D) structure of Mu Ner is essential in order to gain insight into its mode of DNA binding.

Results: The 3D solution structure of Mu Ner has been solved by 3D and 4D heteronuclear magnetic resonance spectroscopy. The structure consists of five α helices, two of which comprise a helix-turn-helix (HTH) motif. Analysis of line broadening and disappearance of cross-peaks in a ^{1}H-^{15}N correlation spectrum of the Mu Ner–DNA complex suggests that residues in these two helices are most likely to be in contact with the DNA.

Conclusions: Like the functionally analogous cro proteins from phages λ and 434, the Mu Ner protein possesses a HTH DNA recognition motif. The Ner protein from phage D108 and the Nlp protein from *E. coli* are likely to have very similar tertiary structures due to high amino-acid-sequence identity with Mu Ner.

Structure 15 October 1995, **3**:1087–1095

Key words: DNA recognition, helix-turn-helix, Mu Ner, NMR, solution structure

Introduction

The Mu Ner protein regulates the switch between the lysogenic and lytic life cycles of phage Mu [1]. When bound to DNA, Mu Ner down-regulates its own transcription and prevents transcription of the cI repressor which is responsible for maintaining the lysogenic state. Mu Ner represents an interesting system for structural studies for a variety of reasons. Despite the functional similarity there is no significant sequence identity between Mu Ner and the cro proteins of other phages, such as phage λ and 434. As the cro proteins employ the helix-turn-helix (HTH) motif for DNA recognition, it is of interest to explore whether functional similarity invariably coincides with structural similarity with respect to the DNA-binding motif. Moreover, only two proteins have been identified so far that have significant amino-acid-sequence identity with Mu Ner, namely the Ner protein of phage D108, which also regulates the lysogenic/lytic switch (and has 53% amino-acid identity with Mu Ner) [2,3] and the Nlp protein of *Escherichia coli*, which is involved in maltose metabolism [4] (and has 65% amino-acid identity).

Interestingly, despite the high level of amino-acid-sequence identity, the two Ner proteins from phages Mu and D108 do not bind to each other's DNA recognition sites. D108 Ner binds to a GC-rich, 47 base-pair (bp) DNA fragment [5], whereas the Mu Ner binding site is an AT-rich 50 bp fragment [6,7]. This suggests that different amino acids are used for DNA recognition in these two proteins. We recently demonstrated that the primary binding site for Mu Ner is a 30 bp DNA fragment, to which two Mu Ner dimers bind [8]. This is substantially longer than the sites identified for the cro proteins of phages λ and 434, which interact with 17 bp and 14 bp stretches, respectively, as dimers [9,10].

Even though the secondary structure of Mu Ner was previously delineated by 2D NMR techniques [11], no 3D structure was determined at the time, mainly due to difficulties arising from severe resonance degeneracy in the 2D NMR spectra. With the advent of 3D- and 4D-heteronuclear NMR techniques [12–14] and uniformly ^{15}N- and $^{15}N/^{13}C$-labeled protein, it finally became feasible to determine the 3D structure of Mu Ner.

Results and discussion

Resonance assignments

The published resonance assignments of the ^{1}H and ^{15}N chemical shifts of Mu Ner [11] did not include those of residues 1–5 and the side chains of residues 11, 18–22, 25, 28, 30, 37, 38, 41, 49, 56, 58–59, 61–63, 69 and 74. Using 3D experiments and uniformly ^{15}N- or $^{15}N/^{13}C$-labeled protein [12–14] we were able to obtain complete ^{1}H, ^{13}C and ^{15}N resonance assignments, with the exception of the NH and ^{15}N resonances of Ser26.

Determination of the 3D structure of Mu Ner

The 3D structure of Mu Ner was solved by 3D and 4D heteronuclear NMR spectroscopy [12–14]. The majority of NOEs between ^{15}N-attached and ^{13}C-attached

*Corresponding authors.

 ISSN 0969-2126

protons, and between ^{13}C-attached protons, were obtained from 4D $^{15}N/^{13}C$-separated and $^{13}C/^{13}C$-separated NOE spectra, respectively. It is worthwhile noting that even the 3D spectra exhibited extensive spectral overlap, demonstrating that helical proteins generally display poor chemical-shift dispersion. An example of the quality of the data is shown in Figure 1 which shows two ^{1}H-^{13}C planes from the 4D $^{13}C/^{13}C$-separated NOE spectrum.

The final 30 simulated-annealing structures were calculated on the basis of 1546 experimental NMR restraints. These comprised 944 approximate interproton distance, 40 hydrogen bonding distance (for 20 backbone hydrogen bonds), 89 torsion angle (56 ϕ, 27 χ_1 and 6 χ_2), 46 $^{3}J_{HN\alpha}$ coupling constant, 140 carbon chemical shift, and 287 proton chemical shift restraint. Stereospecific assignments were obtained for 19 out of 51 β-methylene groups, the methyl groups of the two valines, and the methyl groups of two of the seven leucines. A summary of the structural statistics is provided in Table 1, and a stereoview of a best-fit superposition of the backbone atoms and ordered side chains for the ensemble of 30 final simulated annealing structures is shown in Figure 2. The N (residues 1–7) and C (residues 67–74) termini are disordered in solution. The rest of the protein (residues 8–66) is well defined, with a precision of 0.36±0.06 Å for the backbone (N, Cα, C, O) atoms, 0.95±0.07 Å for all atoms, and 0.39±0.07 Å for all atoms that do not exhibit conformational disorder. The precision is defined as the average root mean square (rms) difference between the individual simulated-annealing structures and the mean coordinate positions (which were obtained by averaging the coordinates of the individual structures best-fitted to residues 8–66). Analysis of the final structures (residues 8–66) indicates that 79% of the residues lie within the most favoured region of the Ramachandran ϕ,ψ plot and 21% in the additional allowed regions [15].

Structure of Mu Ner

The structure of Mu Ner consists of five α helices: H1 (Arg11–Lys20), H2 (Leu25–Gln31), H3 (Pro36–Asn41), H4 (Pro48–Ala57) and H5 (Pro62–Ile65), with short loops connecting them. Helices H2 and H3 form a HTH motif, characterized by two helices approximately orthogonal to each other [16,17] (Fig. 3a). The HTH motif alone is not a stable domain, and therefore has to be stabilized by other structural elements which can consist of α helices or β sheets. In the case of Mu Ner this supporting scaffold is provided by helices H1, H4 and H5.

Mu Ner has a large number of hydrophobic residues (23 out of 74 amino acids): ten alanines, two valines, seven leucines and four isoleucines. Most of these contribute to the hydrophobic core which consists of Val14, Ile15, Leu18, Leu23, Leu25, Leu28, Leu43, Ile53, Ile54, Ala55, Ala57, Leu58, Pro62 and Ile65. The overall shape of the protein is quite compact and globular. A relatively large number of positively charged residues (six lysines and six arginines) result in a high positive electrostatic potential, principally around helix H2 and a loop between helices H2 and H1 (Fig. 3b). It is likely that these are the areas that interact with DNA (see below).

Comparison with other HTH proteins

Most prokaryotic transcriptional regulators contain a HTH DNA-binding domain. In general, these proteins interact with DNA as dimers and bind in the major groove. Among them are the cro and repressor proteins of phages λ, 434, and P22, which are phage proteins involved in the lysogenic/lytic switch. Repressors maintain the lysogenic state, during which the phage DNA remains integrated in the bacterial chromosome, and the cro proteins activate the switch to, and maintain, the lytic cycle, during which the phage DNA is replicated and progeny phage particles are produced. 3D structures of several repressor and cro proteins have been determined

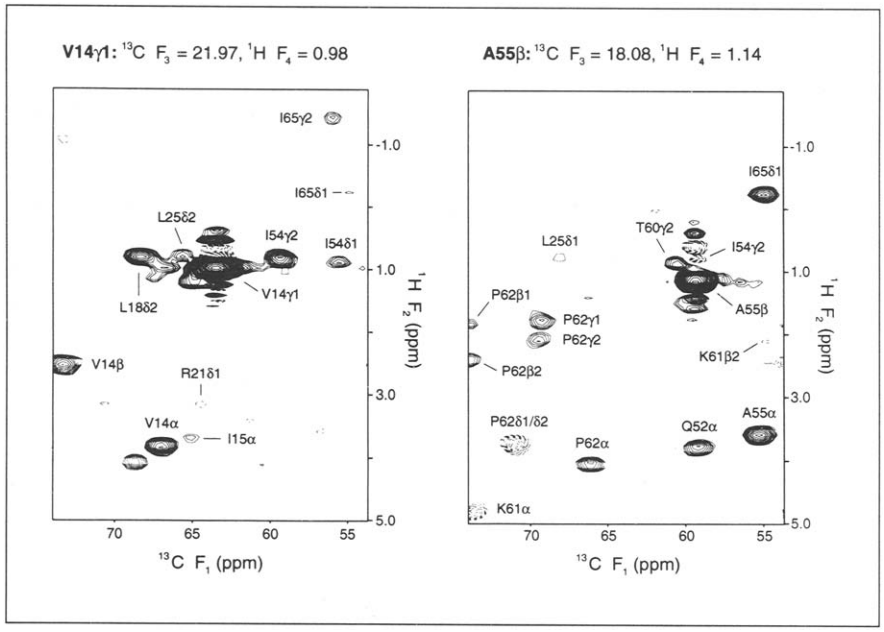

Fig. 1. Two $^{13}C(F_1)$-$^{1}H(F_2)$ planes from the 120 ms mixing time 4D $^{13}C/^{13}C$-separated NOE spectrum of Mu Ner. The destination protons are (left panel) the $C^{\gamma 1}H_3$ methyl group of Val14 and (right panel) the $C^{\beta}H_3$ methyl group of Ala55.

Table 1. Structural statistics*.

	<SA>	$(\overline{SA})_r$
Rms deviations from experimental distance restraints (Å)†		
All (984)	0.026±0.002	0.026
Interresidue sequential (|i - j| = 1) (251)	0.019±0.005	0.011
Interresidue short range (1< |i - j| ≤ 5) (202)	0.038±0.004	0.042
Interresidue long range (|i - j| > 5) (157)	0.036±0.005	0.035
Intraresidue (334)	0.018±0.003	0.016
H-bond (40)‡	0.030±0.007	0.039
Rms deviations from experimental dihedral angle restraints (°) (89)†	0.46±0.07	0.48
Rms deviations from $^3J_{HN\alpha}$ coupling constants (Hz) (46)†	0.61±0.03	0.63
Rms deviations from secondary ^{13}C chemical shifts (ppm)		
$^{13}C\alpha$ (72)	1.07±0.04	1.07
$^{13}C\beta$ (68)	0.92±0.05	0.89
Rms deviations from ^1H chemical shifts (ppm)§		
Overall (287)	0.25±0.005	0.25
CαH (74)	0.26±0.007	0.29
methyl (39)	0.25±0.008	0.26
other (174)	0.24±0.009	0.23
Rms deviations from idealized covalent geometry		
Bonds (Å) (970)	0.005±0.0002	0.006
Angles (°) (1760)	0.629±0.016	0.722
Impropers (°) (514)#	0.433±0.048	0.417
E_{LJ} (kcal mol^{-1})**	−246±5	−234

*<SA> are the final 30 simulated-annealing structures. $\overline{SA}$ is the mean structure obtained by averaging the coordinates of the individual SA structures best-fitted to each other (for residues 8–66), and $(\overline{SA})_r$ is the restrained regularized mean structure obtained from $\overline{SA}$. The number of terms for the various restraints are given in parentheses. †None of the structures exhibit distance violations > 0.4 Å, dihedral-angle violations > 5° or $^3J_{HN\alpha}$ coupling constant violations > 2 Hz. There are no systematic interproton-distance violations between 0.1 Å and 0.4 Å among the 30 structures. ‡For each backbone hydrogen bond there are two distance restraints: $r_{NH\text{-}O}$=1.7–2.5 Å and $r_{N\text{-}O}$=2.3–3.5 Å. §All exchangeable protons (i.e. backbone and side-chain amides) are excluded. #The improper torsion restraints serve to maintain planarity and chirality. **E_{LJ} is the Lennard-Jones van der Waals energy calculated with the CHARMM empirical energy function [69]. It is not included in the target function for simulated annealing.

[15,16]. We shall limit our discussion to those structures that contain five α helices.

The structures of 434 cro [18] and its complex with DNA [19,20] have been determined by X-ray crystallography. A structure of the N-terminal domain of the 434 repressor (residues 1–69) has been determined by X-ray crystallography [21] and NMR spectroscopy [22], and the structure of the 434 repressor–DNA complex was solved by X-ray crystallography [23–26]. Structures of the DNA-binding domain of the λ repressor (residues 1–92) [27] and its complex with operator DNA [28,29] have been determined by X-ray crystallography. Finally, the NMR solution structure of the N-terminal domain of the P22 c2 repressor (residues 1–76) has recently been determined [30].

Amino-acid-sequence comparisons indicate that the cro and repressor proteins from phages λ, 434 and P22 share a common ancestor [31]. For example, the percentage amino-acid-sequence identity between 434 cro and the P22 repressor and between the 434 repressor and the λ

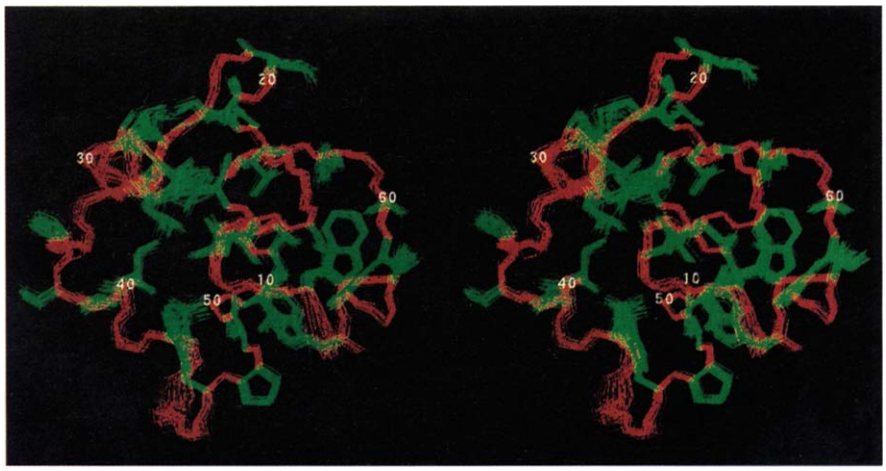

Fig. 2. The best-fit superposition of the backbone and ordered side chains of the 30 final simulated annealing structures of Mu Ner (residues 8–66). The atoms that do not exhibit conformational disorder are: all N, Cα, C, O and Cβ atoms of residues 8–66; the complete side chains of residues Trp9, Val14, Leu18, Leu23, Leu25, Phe32, Pro36, Thr37, Thr38, Leu43, Trp47, Pro48, Ile53, Ile54, Leu58, Thr60, Pro62, Val64, Ile65 and Trp66; the side chains of Asp8, Asp13, Ile15, Leu28, Leu39, Lys49, Glu51 and Gln52 up to Cγ; and the side chains of Lys19, Arg21, Gln31 and Lys61 up to Cδ. The backbone is shown red and side chains in green. (The figure was generated with the program AVS-XPLOR [65].)

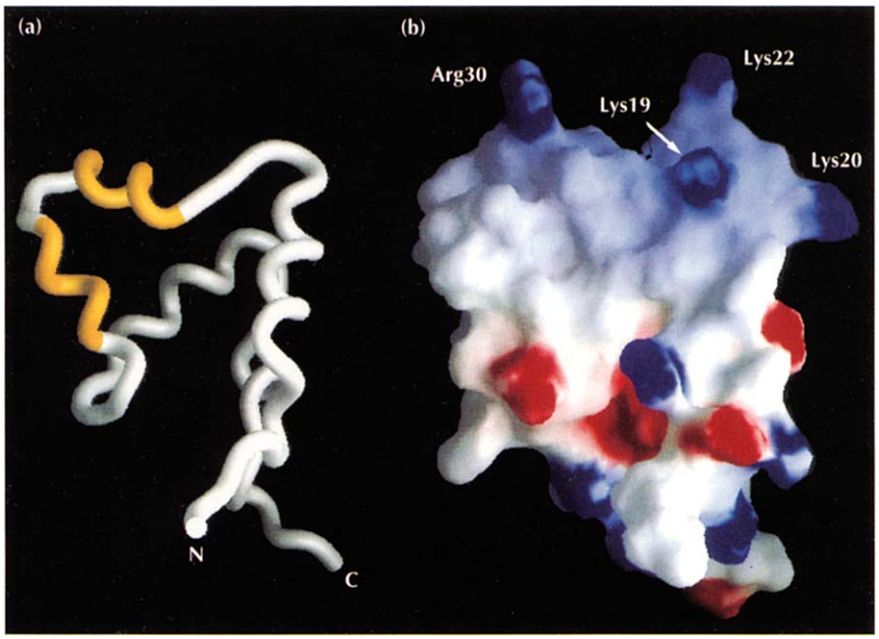

Fig. 3. Overall structure of Mu Ner. **(a)** Backbone fold of Mu Ner, with the HTH motif shown in yellow. **(b)** The surface of Mu Ner color coded by electrostatic potential. Red indicates negative potentials that are less than –7 *kT*, and blue indicates positive potentials that are greater than 7 *kT*. The electrostatic potential was calculated and the images generated using the program GRASP [66]. The labeled residues in (b) indicate residues that give rise to a contiguous patch of positive electrostatic potential.

repressor is 30% and 24%, respectively. In addition, sequence identities for these proteins are highest in the regions containing the HTH motif. In contrast, comparison of the Mu Ner sequence with the sequences of 434 cro and the N termini of the λ, 434 and P22 repressors yields sequence identities of only 11–16% (8 to 12 amino acids). Although, on the basis of the λ repressor structure, it was possible to predict the presence of the HTH motif in the repressor and cro proteins of phages 434 and P22, sequence comparison of Mu Ner with any of these proteins did not yield any convincing alignment and therefore structure prediction [6]. This supports the notion that sequence identities of 10–15% are not sufficient to delineate clear homologies, unless large families of sequences are available.

Figure 4 shows backbone ribbon drawings of 434 cro, the N-terminal domains of the λ, 434 and P22 repressors, and Mu Ner, all in the same orientation. This figure demonstrates, in a striking manner, that despite the very low amino-acid-sequence identity between Mu Ner and the other four proteins, the tertiary folds are almost identical. This is confirmed by a best-fit superpositions of Mu Ner with these four proteins, which yield atomic rms differences of 2.1–2.3 Å for 44 to 48 Cα atoms. A sequence alignment carried out using these structural homologies is provided in Figure 5. Most of the structural similarities are clustered within the helical regions of these proteins. The 'recognition' helix, H3 of Mu Ner, (which provides most of the base-specific contacts) is three amino acids shorter than the corresponding helices of 434 cro and the 434 repressor, two amino acids shorter than H3 of the λ repressor and one amino acid shorter than H3 in the P22 repressor. This difference in length, however, does not affect the position of the N terminus of H3 and therefore the preceding turn. In particular, the

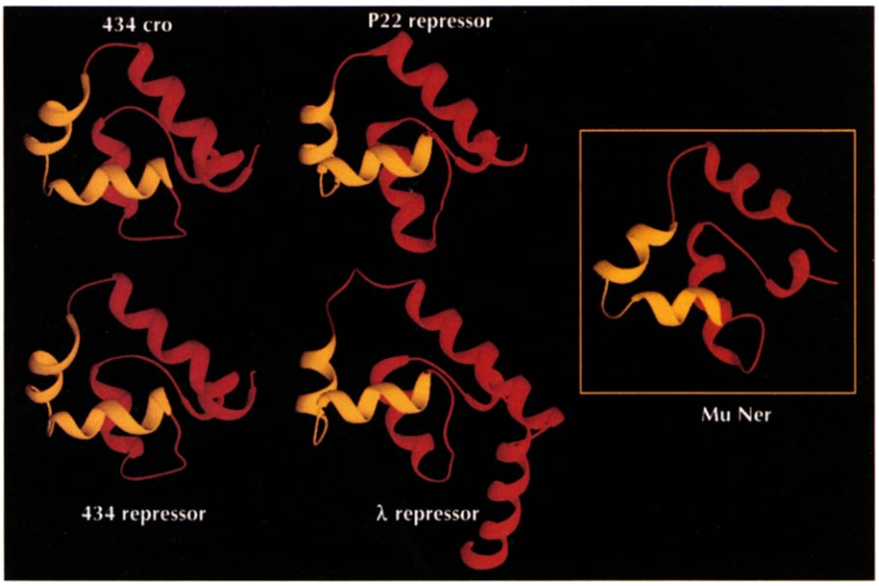

Fig. 4. Ribbon drawings of 434 cro, the 434 repressor (residues 1–69), the P22 repressor (residues 6–65), the λ repressor (residues 1–92) and Mu Ner (residues 8–66), all in the same orientation. The HTH motif is shown in yellow. (The figure was generated using with the program RIBBONS [67].)

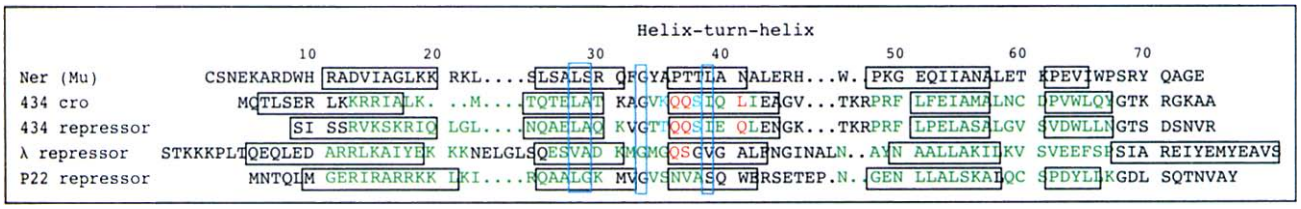

Fig. 5. Sequence alignment, on the basis of structural homology, of Mu Ner, 434 cro, 434 repressor (residues 1–69), λ repressor (1–92) and P22 repressor (1–76). Regions that are structurally homologous to Mu Ner are shown in green, and the helical regions are identified by horizontal boxes. Amino acids shown in red and blue in the 434 cro, 434 repressor and λ repressor sequences participate in specific DNA recognition. Vertical boxes enclose residues which are conserved in the HTH motifs. The structural superpositions with Mu Ner (performed with the program O [68]) were as follows: 434 cro, atomic rms of 2.3 Å for 44 Cα atoms; 434 repressor, atomic rms of 2.3 Å for 48 Cα atoms; λ repressor, atomic rms of 2.1 Å for 44 Cα atoms; and P22 repressor, atomic rms of 2.2 Å for 46 Cα atoms.

important glycine residue between H2 and H3 is conserved, thereby permitting optimal positioning of the recognition helix in most HTH proteins. It is also worth noting that the Cα rms differences between the residues of the HTH motif of Mu Ner (25–41) and these four proteins are: 0.67 Å for the 434 cro (residues 17–33); 0.82 Å for the 434 repressor (17–33); 0.84 Å for the λ repressor (33–49); and 0.72 Å for P22 repressor (21–37).

In the HTH motif, residues 4 and 15 (numbering from the first residue of H2 in Mu Ner) cannot be charged, as they are buried within the protein interior. In the case of Mu Ner, these residues are Leu28 and Leu43, and they participate in the formation of the hydrophobic core. Residue 9 of the motif is usually a glycine, as it is placed in a turn, and in Mu Ner Gly33 occupies this position. Residue 5 is usually small, necessitated by the fact that it is wedged between two α helices, and is frequently an alanine or glycine. In Mu Ner this position is occupied by Ser29. From a structural point of view, therefore, the Mu Ner sequence in the region of helices H2 and H3 and the turn between them contains the correct amino acids for the formation of an HTH motif.

In addition, the positions of several residues from H1, H4 and H5 that contribute to the hydrophobic core of Mu Ner correspond to positions of amino acids of the same type in the other four phage proteins (Fig. 5). In particular, the position corresponding to Val14 in Mu Ner is occupied by hydrophobic residues in three of the four proteins. Within H4, the position corresponding to Ile54 is occupied by either leucine or isoleucine, and the position corresponding to Ala57 is an alanine in three of the proteins. The position corresponding to Leu58 is occupied by leucines in all four proteins. Finally, within H5, the position occupied by Pro62 is either a proline or a valine in the other proteins, and the position corresponding to Ile65 is a leucine in three cases and a proline in one case. Thus, the amino acids that are important in stabilizing the overall fold are clearly conserved between Mu Ner and the other four phage proteins.

Even though phages Mu and λ belong to two different families of phages, Mu belonging to a family of transposable elements and λ (as well as phages 434 and P22) to a family of temperate phages, DNA sequence analysis has indicated that these two phages may have a common ancestor that possessed some common genes, such as the lysis and morphogenesis genes [32]. The structural similarity between Mu Ner and the cro and repressor proteins from phages λ, 434 and P22 provides further support for a common origin for these phages.

Comparison of the amino acid sequence of Mu Ner with the sequences of D108 Ner and the Nlp protein of *E. coli* (Fig. 6), aligned to maximize sequence identity, strongly suggests that the latter proteins probably adopt 3D structures that are very similar to that of Mu Ner. In particular, sequences of Mu Ner, D108 Ner and Nlp are practically identical in the regions corresponding to H1, H4 and H5 of Mu Ner; there is more variation in the regions corresponding to H2 and H3. For H3 this observation can be rationalized on the basis of the different DNA sequences recognized by Mu Ner and D108 Ner, which would result in different amino acids occupying positions 1–3 and/or 5–6 of H3. The DNA-binding recognition sequence of the Nlp protein is not known, but comparison of the putative H3 residues of Nlp and D108 Ner indicates that these two proteins may bind to similar DNA sequences.

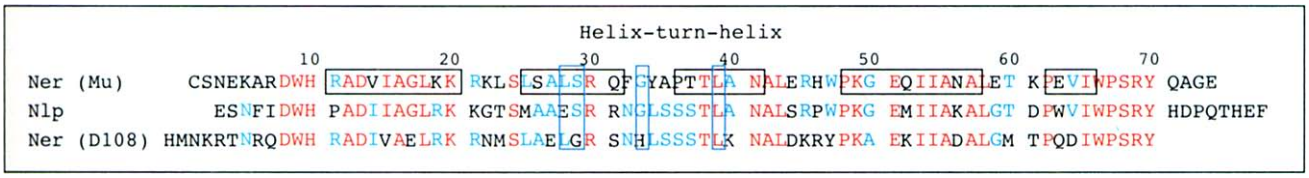

Fig. 6. Sequence alignment of Mu Ner with the Nlp protein of *E. coli* (which shows 65% sequence identity) and the Ner protein from phage D108 (53% sequence identity). Amino acids that are identical in all three proteins are shown in red, and those identical between two proteins in blue. α-helical regions of Mu Ner are identified by horizontal boxes. Vertical boxes enclose residues that are conserved between the HTH motifs.

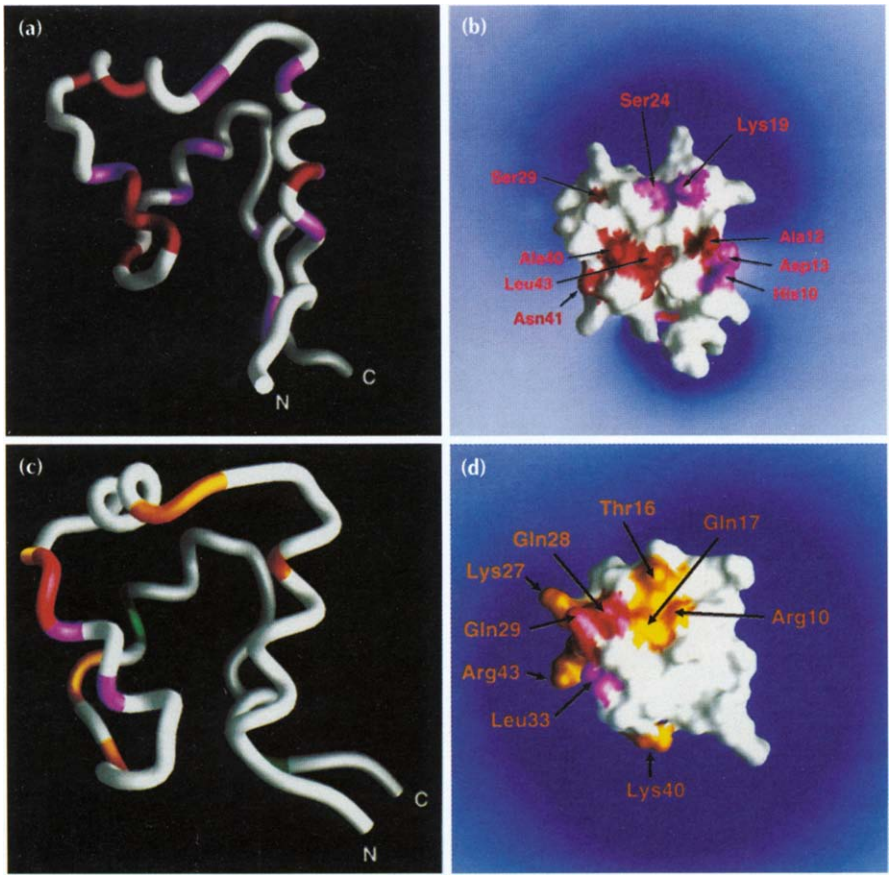

Fig. 7. DNA-binding surface of Mu Ner and 434 cro. **(a,b)** Residues with broadened (magenta) or missing (red) ^{1}H-^{15}N cross peaks in the ^{1}H-^{15}N HSQC spectrum of the complex of Mu Ner with a 16 bp duplex oligonucleotide. ((a) shows the backbone and (b) the solvent-accessible surface of Ner. The blue contours in (b) enclose regions of positive electrostatic potential at 2 kT). **(c)** Residues of 434 cro participating in protein–DNA and protein–protein interactions. The color coding is as follows: red, residues (Gln28 and Gln29) that form hydrogen bonds with the DNA bases; magenta, residues in van der Waals contact with the bases (Ser30, Leu33, also Gln29); orange, residues (Arg10, Thr16, Gln17, Lys27, Lys40, Arg41, Arg43 and Phe44) that form hydrogen bonds with the DNA phosphates; and green, residues at the protein–protein interface in the dimer (Phe46, Glu47, Tyr61, and Arg43). **(d)** DNA-contacting residues mapped onto the solvent-accessible surface of 434 cro. The blue contour encloses a region of positive electrostatic potential at 2 kT. (Potentials were calculated and the figures generated with the program GRASP [67].)

In terms of the structurally conserved features of the HTH motif in Nlp and D108 Ner, each has one amino acid that does not conform to the general rules. Nlp has a glutamate in position 4 of H2 that, in general, is occupied by a hydrophobic residue, and D108 Ner has a histidine in the position within the turn usually occupied by a glycine. It is therefore likely that some of the structural details for these two proteins will be different, but that the overall fold will be virtually identical.

A putative model for Mu Ner–DNA interactions

The structural homology of Mu Ner with the λ and 434 proteins enables one to tentatively suggest which amino acids in Mu Ner are likely to interact with DNA. Residues from the N terminus of H2 and the loop between H2 and H3 interact with the DNA backbone on one side of the major groove, and residues from the loop between H3 and H4 interact with the backbone on the other side of the major groove. These interactions place H3 in the appropriate position within the major groove of the DNA for specific protein–DNA recognition. In the HTH motif of the λ and 434 proteins most of the base-specific recognition is provided by residues −1, 1–3 and 6 of H3. In Mu Ner these positions correspond to Ala35, Pro36, Thr37, Thr38, and Asn41 (Fig 5). All these residues are good candidates for involvement in protein–DNA interactions.

This hypothesis is supported by NMR data of Mu Ner complexed with a double stranded 16 bp DNA fragment comprising the sequence 5′-TAAGCTAGCTAAGTTT-3′ (data not shown). This fragment contains 12 base pairs that overlap with the 30 bp Mu Ner binding site, and Mu Ner binds to it as a dimer [8]. The chemical shifts of amide groups that are involved in protein–DNA interactions undergo large changes upon complex formation. These changes result in signal loss due to line broadening, and can be readily monitored by recording ^{1}H-^{15}N correlation spectra of ^{15}N-labeled protein in the free and bound states [33]. It is worth noting that line broadening, disappearance of cross peaks and/or chemical-shift changes observed in a ^{1}H-^{15}N correlation spectrum upon complex formation provide a well established and reliable indicator of interaction surfaces [33–40].

The following ^{1}H-^{15}N cross peaks were broadened in the ^{1}H-^{15}N correlation spectra of the Mu Ner–DNA complex: His10, Asp13, Gly17, Lys19, Ser24, Thr38, Leu39, Glu51, Ile54, Glu63 and Ser68–Tyr70 (Fig. 7a,b). In addition, the ^{1}H-^{15}N cross peaks of Ala12, Leu28, Ser29, Phe32, Ala40-Glu44 and Trp47 were broadened beyond detection (Fig. 7a,b). Most of the residues whose ^{1}H-^{15}N cross peaks disappeared upon formation of the Mu Ner-DNA complex are clustered within H2 and H3. Figure 7b shows the position of residues with broadened ^{1}H-^{15}N cross peaks mapped onto the solvent-accessible surface of Mu Ner. The main region of positive electrostatic potential surrounds helix H2 and the loop between helices H1 and H2, which would lie in the vicinity of the DNA. Some residues within H1, H4 and H5 also

have broadened ^{1}H-^{15}N cross peaks, and these residues may be involved in protein–protein interactions associated with Mu Ner dimerization on the DNA.

Figures 7c and 7d show the residues of 434 cro that participate in either protein–DNA interactions or protein–protein interactions within the dimer mapped onto the backbone (Fig. 7c) or the solvent-accessible surface of the protein (Fig. 7d). Comparison of Figures 7a and 7b with Figures 7c and 7d shows that, despite similarity in the overall fold between Mu Ner and 434 cro, residues that interact with DNA in Mu Ner appear to be clustered in different areas of the protein surface. In addition, as the overall charge of 434 cro is much higher, practically the whole protein is surrounded by a positive electrostatic field. Clearly, structure determination of a Mu Ner–DNA complex is required to establish the details of the interaction of Mu Ner with DNA.

Biological implications

Mu Ner belongs to a class of proteins, namely the cro and repressor proteins, that regulate the lysogenic/lytic switch in the life cycle of phages. Structures of cro and repressor proteins (N-terminal domains) from phages λ, 434 and P22 have been solved previously. All of these proteins utilize the helix-turn-helix (HTH) motif for DNA recognition and, with the exception of λ cro, consist of five α helices (λ cro has a different topology comprising three α helices and three β strands [41]). Despite the absence of any significant sequence identity between Mu Ner and these other phage proteins, the overall three-dimensional (3D) structure of Mu Ner is very similar to that of 434 cro, and the N-terminal domains of the λ, 434 and P22 repressors. Thus, functional similarity is coupled to structural similarity, lending support to the notion that phages Mu and λ may have a common ancestor. The main difference between Mu Ner and the λ, 434 and P22 proteins is that Mu Ner binds to a 30 bp DNA fragment as a dimer of dimers, whereas the other proteins bind to their DNA target sites as dimers.

Evaluation of the 3D structure of Mu Ner enables us to postulate that two other proteins with very high sequence identity to Mu Ner, namely D108 Ner and the Nlp protein of *Escherichia coli*, also have similar 3D structures.

Materials and methods

Protein expression and purification

Mu Ner was purified from *E. coli* B cells containing the expression plasmid pL-ner, as described previously [42]. Uniformly labeled ^{15}N or $^{15}N/^{13}C$ protein was obtained from bacteria grown in minimal media, with $^{15}NH_4Cl$ and $^{13}C_6$-glucose as the sole nitrogen and carbon sources, respectively. The samples for NMR studies were dialyzed against 50 mM potassium phosphate (pH 7.0), 0.15 M KCl. The following 1 mM NMR samples were used for data collection: ^{15}N-labeled Mu Ner in 90% H_2O/10% D_2O, $^{15}N/^{13}C$-labeled Mu Ner in 90% H_2O/10% D_2O, and $^{15}N/^{13}C$-labeled Mu Ner in 100% D_2O.

All NMR experiments were carried out at 35°C on Bruker AMX500 or AMX600 spectrometers equipped with z-shielded gradient triple resonance probes. The ^{1}H, ^{15}N and ^{13}C chemical shifts of the backbone and side chains were assigned on the basis of through-bond heteronuclear correlations using the following 3D experiments: HNHA, CBCANH, CBCA(CO)NH, HBHA(CO)NH, C(CO)NH, H(CCO)NH, HCCH-COSY and HCCH-TOCSY [12–14]. Approximate interproton distance restraints between NH protons were obtained from a 3D ^{15}N-separated NOE spectrum (mixing time 120 ms) [43]. Distance restraints between NH protons and ^{13}C-attached protons and between ^{13}C-attached protons were obtained from 4D $^{15}N/^{13}C$-separated [44] and $^{13}C/^{13}C$-separated [45,46] NOE spectra, respectively, recorded with mixing times of 120 ms.

Quantitative $^{3}J_{HN\alpha}$, $^{3}J_{CC}$, $^{3}J_{C\gamma N}$, $^{3}J_{C\gamma CO}$ and $^{3}J_{\alpha\beta}$ couplings were measured from a 3D HNHA spectrum [47], a 2D long-range carbon–carbon correlation spectrum [48], a 2D ^{13}C-{^{15}N}-spin-echo difference constant time heteronuclear single-quantum coherence (HSQC) spectrum [49], a 2D ^{13}C-{CO}-spin-echo difference constant time HSQC spectrum [50], and a 3D HACAHB-COSY spectrum, respectively [51]. All NMR spectra were processed with NMRPipe software [52] and analyzed using the programs PIPP, CAPP and STAPP [53].

Structure calculations

Approximate interproton distance restraints derived from the 3D and 4D heteronuclear separated NOE spectra were classified into four ranges: 1.8–2.7 Å (1.8–2.9 Å for NOEs involving NH protons), 1.8–3.3 Å (1.8–3.5 Å for NOEs involving NH protons), 1.8–5 Å and 1.8–6 Å, corresponding to strong, medium, weak and very weak NOEs, respectively [54]. For distances involving methyl protons and non-stereospecifically assigned methylene protons a $(\Sigma r^{-6})^{-1/6}$ sum was employed [55]. In addition, 0.5 Å was added to the upper limit of distances involving methyl protons to account for the higher apparent intensity of methyl resonances [56]. NOEs between protons separated by two bonds, or by three bonds if not stereospecifically assigned, were not incorporated into the restraints set.

ϕ and χ_1 torsion angle restraints were obtained on the basis of the $^{3}J_{HN\alpha}$ and $^{3}J_{\alpha\beta}$ coupling constants combined with intraresidue and sequential interresidue NOEs involving the backbone protons, using the program STEREOSEARCH [57]. The χ_1 angles of both valines, two of the three threonines and all four isoleucines were determined using the $^{3}J_{C\gamma N}$ and $^{3}J_{C\gamma CO}$ coupling constants. The χ_2 angles of two out of the seven leucine residues and of the three isoleucines were obtained from the long-range $^{3}J_{CC}$ coupling constants and the pattern of intraresidue NOEs [58].

The $^{3}J_{HN\alpha}$ coupling constants, which were included directly in the refinement, comprised only those that could be measured from the 3D HNHA experiment to an accuracy of 0.5 Hz or better. The minimum ranges employed for the ϕ, χ_1 and χ_2 torsion angle restraints were ±10°, ±20° and ±20°, respectively [59]. For all residues, the angular standard deviations of the torsion angles for the ensemble of simulated-annealing

structures were much smaller than the ranges employed for the corresponding torsion-angle restraints.

The structures were calculated using the hybrid distance geometry/simulated annealing protocol [60] which makes use of the program X-PLOR (version 3.1) [61] modified to include coupling-constant [62], carbon chemical-shift [63] and proton chemical-shift [64] restraints. The target function that is minimized during simulated annealing contains the following terms (the final values of the force constants for each term are given in parentheses): quadratic harmonic potential terms for covalent geometry, that is bonds (500 kcal mol^{-1} $Å^{-2}$), angles and improper torsions (500 kcal mol^{-1} rad^{-2}); square-well quadratic potentials for the experimental-distance (30 kcal mol^{-1} $Å^{-2}$), and torsion-angle (200 kcal mol^{-1} rad^{-2}) restraints; a quartic van der Waals repulsion term for the non-bonded contacts (4 kcal mol^{-1} $Å^{-4}$); and quadratic harmonic potential terms for the $^3J_{HN\alpha}$ coupling-constant restraints (1 kcal mol^{-1} Hz^{-2}), carbon chemical-shift restraints (0.5 kcal mol^{-1} ppm^{-2}) and proton chemical-shift restraints (7.5 kcal mol^{-1} ppm^{-2}). No hydrogen bonding, electrostatic or 6–12 Lenard–Jones empirical potential terms are present in the target function.

Mu Ner–DNA interactions

The double-stranded 16 bp DNA fragment was synthesized and purified as described previously [8]. The Mu Ner-DNA complexes were formed in either 50 mM potassium phosphate (pH 7.0) and 0.15 M KCl, or in 10 mM potassium phosphate (pH 7.0) in dilute solution and concentrated to about 1 mM using MicroProDiCon device (Spectrum). The 2D ^1H-^{15}N HSQC spectra were recorded at 35°C.

The coordinates of the 30 simulated annealing structures and the restrained regularized mean structure, together with the experimental restraints and complete ^1H, ^{15}N and ^{13}C resonance assignments, have been deposited in the Brookhaven Protein Data Bank.

Acknowledgements: We thank Robb Clubb for useful discussions, Dan Garrett and Frank Delaglio for software support, and Rolf Tschudin for hardware support. This work was supported by the AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health (to GMC and AMG).

References

1. Van de Putte, P., Giphart-Gassler, M., Goosen, N., Goosen, T. & Van Leerdam, E. (1980). Regulation of integration and replication functions of bacteriophage Mu. *Cold Spring Harb. Symp. Quant. Biol.* **45**, 347–353.
2. Hull, R.A., Gill, G.S. & Curtis, R., III (1978). Genetic characterization of Mu-like bacteriophage D108. *J. Virol.* **27**, 513–518.
3. Gill, G.S., Hull, R.C. & Curtis, R., III (1981). Mutator bacteriophage D108 and its DNA: An electron microscopic characterization. *J. Virol.* **37**, 420–430.
4. Choi, Y.-L., Nishida, T., Kawamukai, M., Utsumi, R., Sakai, H. & Komano, T. (1989). Cloning and sequencing of an *Escherichia coli* gene, nlp, highly homologous to the ner genes of bacteriophages Mu and D108. *J. Bacteriol.* **171**, 5222–5225.
5. Kukolj, G., Tolias, P.P., Autexier, C. & DuBow, M.S. (1989). DNA-directed oligomerization of the monomeric Ner repressor from the Mu-like bacteriophage D108. *EMBO J.* **8**, 3141–3148.
6. Tolias, P.P. & DuBow, M.S. (1986). The overproduction and characterization of the bacteriophage Mu regulatory DNA-binding protein ner. *Virology* **148**, 298–311.
7. Kukolj, G., Tolias, P.P. & DuBow, M.S. (1989). Purification and characterization of the Ner repressor of bacteriophage Mu. *FEBS Lett.* **244**, 369–375.
8. Strzelecka, T.E., Hayes, J.J., Clore, G.M. & Gronenborn, A.M. (1995). DNA binding specificity of the Mu Ner protein. *Biochemistry* **34**, 2946–2955.
9. Maniatis, T., *et al.*, & Maurer, R. (1975). Recognition sequences of repressor and polymerase in the operators of of bacteriophage lambda. *Cell* **5**, 109–113.
10. Ptashne, M. (1986). *A Genetic Switch*, Cell Press and Blackwell Scientific Publications, Cambridge, MA.
11. Gronenborn, A.M., Wingfield, P.T. & Clore, G.M. (1989). Determination of the secondary structure of the DNA binding protein Ner from phage Mu using ^1H homonuclear and ^{15}N-^1H heteronuclear NMR spectroscopy. *Biochemistry* **28**, 5081–5089.
12. Clore, G.M. & Gronenborn, A.M. (1991). Structures of larger proteins in solution: three- and four-dimensional heteronuclear NMR spectroscopy. *Science* **252**, 1390–1399.
13. Clore, G.M. & Gronenborn, A.M. (1991). Application of three- and four-dimensional heteronuclear NMR spectroscopy to protein structure determination. *Prog. Nucl. Magn. Res. Spectr.* **23**, 43–92.
14. Bax, A. & Grzesiek, S. (1993). Methodological advances in protein NMR. *Accounts. Chem. Res.* **26**, 131–138.
15. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291.
16. Harrison, S.C. (1991). A structural taxonomy of DNA-binding domains. *Nature* **353**, 715–719.
17. Pabo, C.O. (1992). Transcription factors: structural families and principles of DNA recognition. *Ann. Rev. Biochem.* **61**, 1053–1095.
18. Mondragon, A., Wolberger, C. & Harrison, S.C. (1989). Structure of phage 434 Cro protein at 2.35 Å resolution. *J. Mol. Biol.* **205**, 179–188.
19. Wolberger, C., Dong, Y.-C., Ptashne, M. & Harrison, S.C. (1988). Structure of phage 434cro/DNA complex. *Nature* **335**, 789–795.
20. Mondragon, A. & Harrison, S.C. (1991). The phage 434 cro-OR_1 complex at 2.5 Å resolution. *J. Mol. Biol.* **219**, 321–334.
21. Mondragon, A., Subbiah, S., Almo, S.C., Drottar, M. & Harrison, S. C. (1989). Structure of the amino-terminal domain of phage 434 repressor at 2.0 Å resolution. *J. Mol. Biol.* **205**, 189–201.
22. Neri, D., Billeter, M. & Wüthrich, K. (1992). Determination of the nuclear magnetic resonance solution structure of the DNA-binding domain (residues 1–69) of the 434 repressor and comparison with the X-ray crystal structure. *J. Mol. Biol.* **223**, 743–767.
23. Anderson, J.E., Ptashne, M. & Harrison, S.C. (1987). Structure of the repressor-operator complex of bacteriophage 434. *Nature* **326**, 846–852.
24. Aggarwal, A.K., Rodgers, D.W., Drottar, M., Ptashne, M. & Harrison, S.C. (1988). Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. *Science* **242**, 899–907.
25. Shimon, L.J.W. & Harrison, S.C. (1993). The phage 434 O_R2/R1-69 complex at 2.5 Å resolution. *J. Mol. Biol.* **232**, 826–838.
26. Rodgers, D.W. & Harrison, S.C. (1993). The complex between phage 434 repressor DNA-binding and operator site OR3: structural differences between consensus and non-consensus half-sites. *Structure* **1**, 227–240.
27. Pabo, C.O. & Lewis, M.C. (1982). The operator-binding domain of λ repressor: structure and DNA recognition. *Nature* **298**, 443–447.
28. Jordan, S.R. & Pabo, C.O. (1988). Structure of the λ complex at 2.5 Å resolution: details of the repressor-operator interactions. *Science* **242**, 893–899.
29. Beamer, L. & Pabo, C.O. (1992). Refined 1.8 Å crystal structure of the λ repressor-operator complex. *J. Mol. Biol.* **227**, 177–196.
30. Sevilla-Sierra, P., Otting, G. & Wüthrich, K. (1994). Determination of the nuclear magnetic resonance structure of the DNA-binding domain of the P22 c2 repressor (1 to 76) in solution and comparison with the DNA-binding domain of the 434 repressor. *J. Mol. Biol.* **235**, 1003–1020.
31. Sauer, R.T., Yocum, R.R., Doolittle, R.F., Lewis, M. & Pabo, C.O. (1982). Homology among DNA-binding proteins suggests use of a conserved super-secondary structure. *Nature* **298**, 447–451.
32. Kamp, D. (1987). The evolution of Mu. In *Phage Mu* (Symonds, N., Toussaint, A., van der Putte, P. & Howe, M.M., eds), pp. 259–269, Cold Spring Harbor Laboratory, USA.
33. Dekker, N., Cox, M., Boelens, R., Verrijzer, C.P., van der Vliet, P.C. & Kaptein, R. (1993). Solution structure of the POU-specific DNA-binding domain of Oct-1. *Nature* **362**, 852–855.
34. Gronenborn, A.M. & Clore, G.M. (1993). Identification of the contact surface of a Streptococcal protein G domain complexed with a human Fc fragment. *J. Mol. Biol.* **233**, 331–335.
35. Chen, Y., Reizer, J., Saier, M.H., Fairbrother, W.J. & Wright, P.E. (1993). Mapping of the binding interfaces of proteins of the bacterial phosphotransferase system HPr and IIA^{Glc}. *Biochemistry* **32**, 32–37.
36. van Nuland, N.A., *et al.*, & Robillard, G.T. (1993). The NMR determination of the IIAmtl binding site on HPr of the *Escherichia coli* phosphoenol pyruvate-dependent phosphotransferase system. *FEBS Lett.* **315**, 11–15.

37. Baumann, H., Knapp, S., Karshikoff, A., Ladenstein, R. & Härd, T. (1995). DNA-binding surface of the Sso7d protein from *Sulfolobus solfataricus*. *J. Mol. Biol.* **247**, 840–846.
38. Wikström, M., Sjöbring, U., Drakenberg, T., Forsen, S. & Björk, L. (1995). Mapping of the immunoglobulin light-chain-binding site of protein L. *J. Mol. Biol.* **250**, 128–133.
39. Clubb, R.T., Omichinski, J.G., Savilahti, H., Mizuuchi, K., Gronenborn, A.M. & Clore, G.M. (1994). A novel class of winged helix-turn-helix protein: the DNA-binding domin of Mu transposase. *Structure* **2**, 1041–1048.
40. Kato, K., *et al.*, & Roberts, G.C.K. (1995). Model for the complex between protein G and an antibody Fc fragment in solution. *Structure* **3**, 79–85.
41. Anderson, W.F., Ohlendorf, D.H., Takeda, Y. & Matthews, B. (1981). Structure of the cro repressor from bacteriophage λ and its interaction with DNA. *Nature* **290**, 754–758.
42. Allet, B., Payton, M., Mattaliano, R.J., Gronenborn, A.M., Clore, G. M. & Wingfield, P.T. (1988). Purification and characterization of the DNA binding protein Ner of bacteriophage Mu. *Gene* **65**, 259–268.
43. Marion, D., *et al.*, & Clore, G.M. (1989). Overcoming the overlap in the assignment of ^{1}H-NMR spectra of larger proteins using three-dimensional heteronuclear ^{1}H-^{15}N Hartmann-Hahn and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin-1β. *Biochemistry* **26**, 6150–6156.
44. Kay, L.E., Clore, G.M. & Gronenborn, A.M. (1990). Four-dimensional heteronuclear triple resonance NMR spectroscopy of interleukin-1β in solution. *Science* **249**, 411–414.
45. Clore, G.M., Kay, L.E., Bax, A. & Gronenborn A.M. (1991). Four-dimensional ^{13}C/^{13}C-edited nuclear Overhauser enhancement spectroscopy of a protein in solution: application to interleukin-1β. *Biochemistry* **30**, 12–18.
46. Vuister, G.W., *et al.*, & Bax, A. (1993). Increased resolution and improved spectral quality in four-dimensional ^{13}C/^{13}C-separated HMQC-NOE-HMQC spectra using pulsed field gradients. *J. Magn. Res. B* **101**, 210–213.
47. Vuister, G. W. & Bax, A. (1993). Quantitative *J* correlation: A new approach for measuring homonuclear three-bond $J(H^{N}H^{\alpha})$ coupling constants in ^{15}N-enriched proteins. *J. Am. Chem. Soc.* **115**, 7772–7777.
48. Bax, A., Max, D. & Zax, D. (1992). Measurement of long-range ^{13}C-^{13}C *J* couplings in a 20 kDa protein–peptide complex. *J. Am. Chem. Soc.* **114**, 6923–6925.
49. Vuister, G.W., Wang, A.C. & Bax, A. (1993). Measurement of three-bond nitrogen-carbon *J* couplings in proteins uniformly enriched in ^{15}N and ^{13}C. *J. Am. Chem. Soc.* **115**, 5334–5335.
50. Grzesiek, S., Vuister, G.W. & Bax, A. (1993). A simple and sensitive experiment for measurement of J_{CC} couplings between backbone carbonyl and methyl carbons in isotopically enriched proteins. *J. Biomol. NMR.* **3**, 487–493.
51. Grzesiek, S., Kuboniwa, H., Hinck, A. P. & Bax, A. (1995). Multiple-quantum line narrowing for measurement of H^{α}-H^{β} *J* coupling in isotopically enriched proteins. *J. Am. Chem. Soc.* **117**, 5312–5315.
52. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, J., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* in press.
53. Garrett, D.S., Powers, R., Gronenborn, A.M. & Clore, G.M. (1991). A common sense approach to peak picking two-, three-, and four-dimensional spectra using automatic computer analysis of contour diagrams. *J. Magn. Res.* **95**, 214–222.
54. Clore, G.M., Nilges, M., Sukumaran, D.K., Brünger, A.T., Karplus, M. & Gronenborn, A.M. (1986). The three-dimensional structure of a1-purothionin in solution: combined use of nuclear magnetic resonance, distance geometry and restrained molecular dynamics. *EMBO J.* **5**, 2729–2735.
55. Nilges, M. (1993). A calculational strategy for for the structure determination of symmetric dimers by ^{1}H NMR. *Proteins* **17**, 295-309.
56. Clore, G.M., Gronenborn, A.M., Nilges, M. & Ryan, C.A. (1987). The three-dimensional structure of potato carboxypeptidase inhibitor in solution: a study using nuclear magnetic resonance, distance geometry and restrained molecular dynamics. *Biochemistry* **26**, 8012–8023.
57. Nilges, M., Clore, G.M. & Gronenborn, A.M. (1990). ^{1}H NMR stereospecific assignments by conformational data-base searches. *Biopolymers* **29**, 813–822.
58. Powers, R., Garrett, D.S., March, C.J., Frieden, E.A., Gronenborn, A. M. & Clore, G.M. (1993). The high resolution three-dimensional solution structure of human interleukin-4 determined by multidimensional heteronuclear magnetic resonance spectroscopy. *Biochemistry* **32**, 6744–6762.
59. Lodi, P.J., *et al.*, & Clore, G.M. (1994). High resolution solution structure of the β chemokine hMIP-1β by multi-dimensional NMR. *Science* **263**, 1762–1767.
60. Nilges, M., Clore, G.M. & Gronenborn, A.M. (1988). Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS Lett.* **229**, 317–324.
61. Brünger, A.T. (1993). *X-PLOR Version 3.1 Manual*, Yale University, New Haven, CT.
62. Garrett, D.S., *et al.*, & Clore, G.M. (1994). The impact of direct refinement against three-bond HN-CαH coupling constants on protein structure determination by NMR. *J. Magn. Res. B* **104**, 99–103.
63. Kuszewski, J., Qin, J., Gronenborn, A.M. & Clore, G.M. (1995). The impact of direct refinement against $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts on protein structure determination by NMR. *J. Magn. Res. B* **106**, 92–96.
64. Kuszewski, J., Gronenborn, A.M. & Clore, G.M. (1995). The impact of direct refinement against proton chemical shifts on protein structure determination by NMR. *J. Magn. Res. B* **107**, 293–297.
65. Delano, W. & Brünger, A.T. (1993). *AVS-XPLOR User's Manual*, Yale University, New Haven, CT.
66. Nicholls, A., Sharp, K.A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296.
67. Carson, M. (1987). Ribbon models of macromolecules. *J. Mol. Graphics* **5**,103–106.
68. Jones, T.A. & Kjeldgaard, M. (1992). *O-The manual. Version 5.8.1*, University of Uppsala, Sweden.
69. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. & Karplus, M. (1983). CHARMM: a program for macromolecular energy minimization and dynamics calculations. *J. Comput. Chem.* **4**, 187–217.

Received: **17 Jul 1995**; revisions requested: **11 Aug 1995**; revisions received: **16 Aug 1995**. Accepted: **16 Aug 1995.**